Precise control of neural activity using temporally dithered and spatially multiplexed electrical stimulation

Nishal P. Shah[1,2,4], Sasi Madugula[2,4], A.J. Phillips[1,4], Lauren Grosberg[2,4], Alex R. Gogliettino[4], Aditya Dusi[1] , Pulkit Tandon[1], Jeff Brown[1], Pawel Hottowy[6], Wladyslaw Dabrowski[6], Alexander Sher[7], Alan M. Litke[7], Subhasish Mitra[1], E.J. Chichilnisky[2,3,4]

[1]Department of Electrical Engineering, [2]Department of Neurosurgery, [3]Department of Ophthalmology, [4]Hansen Experimental Physics Laboratory, Stanford University, Stanford CA 94305, USA. [6]AGH University of Science and Technology, Faculty of Physics and Applied Computer Science, 30-059 Krakow, Poland. [7]Santa Cruz Institute for Particle Physics, University of California, Santa Cruz, CA 94064, USA

Abstract

Electrical stimulation with neural implants can restore lost sensory function by evoking patterns of activity in neural populations. However, stimulation with many electrodes generally combines nonlinearly to influence neural activity, and is thus difficult to control. To overcome this challenge, we propose a dynamic stimulation approach that exploits the slow time scales of downstream neural processing and the independence of distant electrodes by encoding a complex visual stimulus into a rapid, greedily chosen, temporally dithered and spatially multiplexed sequence of simple stimulation patterns. The approach was evaluated using a lab prototype of a retinal implant: large-scale, high-resolution multi-electrode stimulation and recording of primate retinal ganglion cells *ex vivo*. Greedy dithering and multiplexing provided a powerful framework for optimizing electrical stimulation, greatly enhancing expected performance compared to existing open loop approaches. The modular framework enabled parallel extensions to naturalistic and dynamic viewing conditions, optimization of perceptual similarity measures and efficient hardware implementation for retinal implants.

Introduction

A major goal of sensory neuroscience is to exploit our understanding of neural circuits to develop implantable devices that can artificially control neural activity for restoring senses such as vision (Humayun et al. 2012; Stingl et al. 2013; Palanker et al. 2020; Beauchamp et al. 2020), audition (Gaylor et al. 2013) and somatosensation (Johnson et al. 2013; Flesher et al. 2016; Armenta Salas et al. 2018). Recent innovations in large-scale and high-resolution electrical stimulation hardware hold great promise for such applications. However, the effect of stimulation with many closely spaced electrodes on neural activity is generally complex and nonlinear, which severely limits the impact of hardware innovations on producing desired spatio-temporal patterns of neural activity.

This paper presents a novel approach to this problem in the context of an epiretinal implant for restoring vision in people blinded by photoreceptor degenerative diseases (Humayun et al. 2012; Beyeler et al. 2019; Bloch and da Cruz 2020). Epiretinal implants electrically activate retinal ganglion cells (RGCs) that have survived degeneration, causing them to send artificial visual signals to the brain. After initial successes, progress towards restoring high-fidelity natural vision using this approach has slowed, likely in part due to indiscriminate activation of many RGCs of different cell types and a resulting inaccurate neural representation of the target stimulus. One reason for this indiscriminate activation is the difficulty of predicting the neural activity evoked by multi-electrode stimulation based on the activity evoked by single-electrode stimulation.

We present a way to bypass this problem by dynamically combining simpler stimulation patterns (Beauchamp et al. 2020) using a novel temporal dithering and spatial multiplexing approach. The presented solution is divided into three steps, allowing it to be modified for a wide range of neural systems and implants. First, we develop a simple, explicit model of how the visual image could be reconstructed in the brain from the activity of many RGCs of diverse types. Second, we avoid the complexity of non-linear electrical stimulation by empirically calibrating RGC responses to a collection of simple single-electrode stimuli which can then be combined asynchronously and sparsely to reproduce patterns of neural activity. Finally, we optimize visual scene reconstruction by greedily selecting a sequence of these simple stimuli, temporally dithered to exploit the high speed of electrically evoked neural responses, and spatially multiplexed to exploit the independence of distant electrodes. These three steps result in a stimulation paradigm in which a visual stimulus is converted into a spatiotemporal pattern of electrical stimuli designed to produce the desired pattern of neural activity.

The approach is tested and evaluated using large-scale multielectrode stimulation and recording from primate retina *ex vivo*, a lab prototype for a future implantable system. The greedy dithering and multiplexing method produced substantial improvements in expected stimulus reconstruction compared to existing open loop methods, with performance approaching that of the best possible algorithm. The algorithm was useful in identifying a subset of the most useful electrodes for a particular retina, which could substantially reduce power consumption in an implant. Extensions of the algorithm can in principle be used to translate the approach to naturalistic viewing conditions with eye movements, and to exploit perceptual metrics to enhance the quality of reconstruction.

Results

First, we frame the translation of a visual stimulus into electrical stimulation as an optimization problem and present a greedy temporal dithering algorithm to solve it efficiently. Then, we use data from the *ex vivo* lab prototype to evaluate the performance of the approach. We compare it with existing methods and develop extensions for spatial multiplexing, natural viewing with eye movements and perceptual quality measures.

The *ex vivo* lab prototype consists of electrical recording and stimulation of the primate retina with a large-scale high-density multi-electrode array (512 electrodes, 60 μm spacing). The visual and electrical response properties of recorded cells are estimated using experimental methods described previously (see Methods). These data provide reliable experimental access to complete populations of ON and OFF parasol cell types, so these two cell types are the focus of the empirical analysis.

Greedy temporal dithering to replicate neural code

Converting a visual stimulus into effective electrical stimulation can be framed as an optimization problem. Using the terminology of optimization, the three key components are the *objective function*, the *constraints* and the *algorithm* (Figure 1A). The *objective* function to be minimized is identified as the difference between the target visual stimulus and a reconstruction of the stimulus from the neural responses, as a proxy for how the brain could use the signal for visual inference. However, certain *constraints* are imposed by electrical stimulation, which provides imperfect control over the activity of a population of cells. Hence, the optimization a*lgorithm* must convert incoming visual stimuli into electrical stimuli, such that the stimulus reconstructed from electrically-evoked responses matches the true stimulus as closely as possible.

Objective: Reconstructing the visual stimulus from neural responses

The *objective* of electrical stimulation in this context is to reproduce, as closely as possible, a visual sensation that would be produced by normal light-evoked responses. However, it is not immediately clear how the brain interprets RGC light responses, and thus how to frame the problem computationally. As a simple proxy, the objective is defined as reconstructing the visual image as accurately as possible from RGC inputs (see Discussion). For simplicity, linear reconstruction is assumed, with parameters adjusted to minimize the mean squared error between the target and the reconstructed stimulus.

Specifically, for a target image shown to the retina in the lab prototype, the reconstructed stimulus is modeled as the linear superposition of spatial filters, each associated with a particular RGC, and weighted by the corresponding RGC response (Figure 1B, (Warland, Reinagel, and Meister 1997; Brackbill et al. 2020)). The reconstruction filter for each ON and OFF parasol cell was estimated using the measured spatial receptive field of the cell obtained with white noise stimulation (Brackbill et al. 2020), scaled to predict the average spike count recorded within 50 ms of the onset of a flashed checkerboard stimulus. Note that in an implanted blind retina, this filter would have to be estimated using spontaneous neural activity instead (see Discussion).

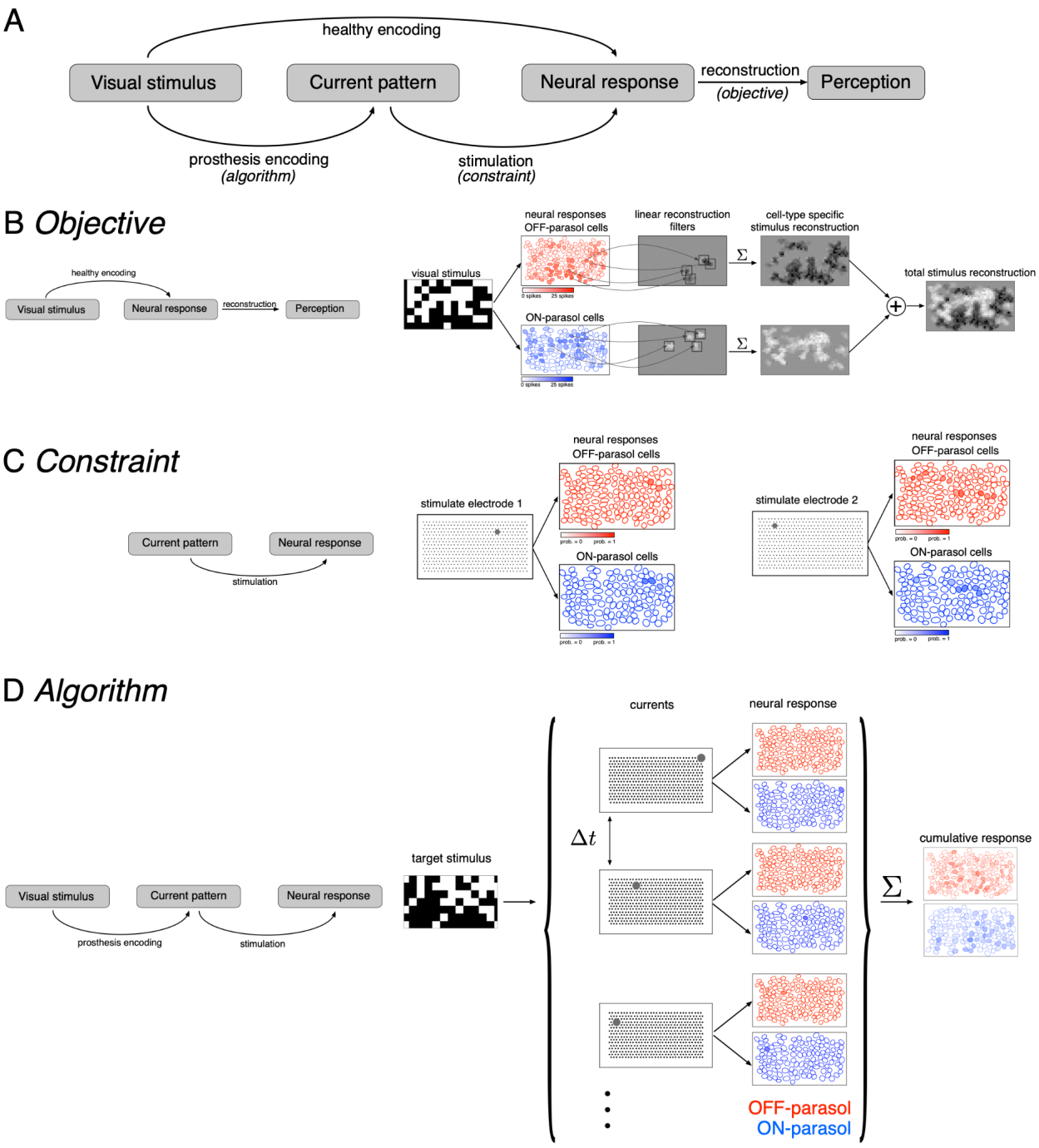

Figure 1: **Algorithmic components of the proposed framework for electrical stimulation.** (A) In a healthy retina, the visual stimulus is encoded in the neural response pattern of retinal ganglion cells (top row). In a retina with an implant, the visual stimulus is encoded into current patterns, which generate neural response patterns (bottom row). In either case, the neural responses are eventually processed by the brain to elicit perception; through a process assumed to involve reconstruction of the image. Selecting the appropriate electrical stimulation can be framed as an optimization problem, in which the goal is to identify an *algorithm* (prosthesis encoding) that achieves an *objective* (reconstruction error) while operating under *constraints* (electrical stimulation). (B) *Objective:* Linear reconstruction of visual stimulus by summing cell-specific spatial filters, weighted by spike counts. Receptive fields of ON (blue) and OFF

(red) parasol cells in a population are shown. (C) *Constraint:* Characterizing electrically evoked RGC responses with a dictionary of stimulation patterns. Example dictionary elements, with cells shaded according to evoked response probability. A single electrode stimulated multiple cells, indicating poor selectivity. (D) *Algorithm:* Run-time usage of the artificial retina. Exploiting the slow visual integration time, distant electrodes are stimulated in fast sequence. The resulting neural response is the summation of spikes elicited in each time step.

Constraint: Collection of neural responses that can be electrically evoked

The limited precision of electrical stimulation *constrains* our ability to produce desired response patterns in the neural population. To incorporate this constraint in the optimization of stimulation, it would be ideal to have a model that characterizes how neurons would respond to arbitrary electrical stimulus patterns produced with the electrode array. Unfortunately, estimating this model is difficult due to nonlinear interactions in neural activation by current passed simultaneously through multiple electrodes (Jepson et al. 2014). An alternative to modeling these complex and high-dimensional nonlinearities is to use a *dictionary* of simple current patterns, for which the activation probability of each cell is calibrated empirically in advance (Figure 1C). For data collected using the lab prototype, a dictionary of single-electrode stimuli is used, in which current is passed through each of the 512 electrodes at each of 40 current levels (logarithmically spaced over the range 0.1-4 μA) and the response probability for each cell is estimated using the fraction of trials in which the cell generated a spike. The evoked spikes were identified using a custom spike sorting algorithm that estimated the electrical artifact produced by stimulation, and matched the residual recorded voltage to template waveforms of cells previously identified during visual stimulation (Mena et al. 2017). In general, even though some stimuli selectively activated one or a few cells (Figure 1C), many stimuli simultaneously activated greater numbers of cells, largely due to axonal stimulation (Figure 1C, Grossberg et al., 2017). High amplitude stimuli that led to the activation of cells with receptive fields off the electrode array were detected by their bi-directional propagation to the edge of the electrode array and were removed from the dictionary (Grosberg et al. 2017; Tandon et al. 2021).

Algorithm: Greedy temporal dithering

Because none of the individual, calibrated stimuli is likely to create a desired pattern of activity across the population, multiple stimuli must be combined, while also avoiding the non-linear interactions mentioned above. This is achieved by rapid *temporal dithering* of a diverse collection of stimuli. If these stimuli are provided in rapid succession, faster (e.g. 0.1 ms interval) than a typical visual integration time (e.g. 50 ms) (see Discussion), the brain presumably cannot distinguish sequential stimulation from simultaneous stimulation, and the evoked perception depends only on the total evoked spike count (Figure 1D). Concretely, the overall problem reduces to finding a sequence of dictionary elements $\{C_1, \cdots, C_T\}$ that minimizes the expected mean squared error between the target visual stimulus and reconstructed responses based on the total spike count:

$$C_1, \ldots, C_T = argmin \; E_{R_i \sim Bernoulli(p_{C_i})} || S - A(R_1 + \ldots + R_T) ||^2 \quad (1)$$

where S is the target visual stimulus, A is the stimulus reconstruction filter and R_i is a vector of the spike counts in the population of cells generated by stimulation C_i, with spikes being drawn according to a Bernoulli process with probability p_i.

A simple, real-time algorithm to produce an effective stimulation sequence is to greedily minimize the expected error between the target image and the resulting reconstruction. This greedy approach makes possible an efficient implementation on low-power hardware. A crucial assumption of the algorithm is that the total spike count evoked by successive stimuli is the sum of the spikes generated by each stimulus. However, when a cell is repeatedly stimulated, the activation probabilities associated with later stimuli – when the cell is in its refractory period – is reduced. To avoid this non-independence, the stimulus at a given moment is chosen from a 'valid' subset of the dictionary that does not include cells that were targeted recently (see Methods).

Greedy temporal dithering outperforms open loop methods

The greedy dithering approach was evaluated on data from the lab prototype. For random checkerboard visual stimulus targets, the greedy dithering stimulation sequence was calculated, neural responses were sampled using measured response probabilities evoked by the individual selected stimuli, and then the target image was linearly reconstructed from these responses. During the stimulation sequence, the reconstructed image slowly built up to a spatially smooth version of the original target image. Not surprisingly, ON (OFF) parasol cells were stimulated more than OFF (ON) parasol cells in bright (dark) regions of the target stimulus (Figure 2). Moreover, the reconstruction for individual trials was similar to the average across multiple trials (Figure 2), indicating that the noise from inter-trial response variation was relatively small (not shown).

To quantify the performance of the greedy temporal dithering approach, its reconstruction error was compared with the error of alternate approaches. First, dithering was compared with open loop stimulation that does not use calibrated electrical and visual response properties to optimize stimulation, as is the case in all existing retinal implants, which instead map the intensity of visual stimulus near each electrode to its stimulation current amplitude (Humayun et al. 2012; Stingl et al. 2013; Palanker et al. 2020). Open loop performance was simulated with the high-density lab prototype electrode array for a direct comparison to greedy dithering. The current through each electrode was given by a sigmoidal function of the intensity of the visual stimulus near the electrode, and the sigmoids across all electrodes were jointly optimized for minimizing the error between the target and its reconstruction. Even with this flexibility,

open-loop control resulted in significantly less accurate reconstruction of the target image (Figure 3D,H), likely due to coactivation of overlapping ON and OFF cell types.

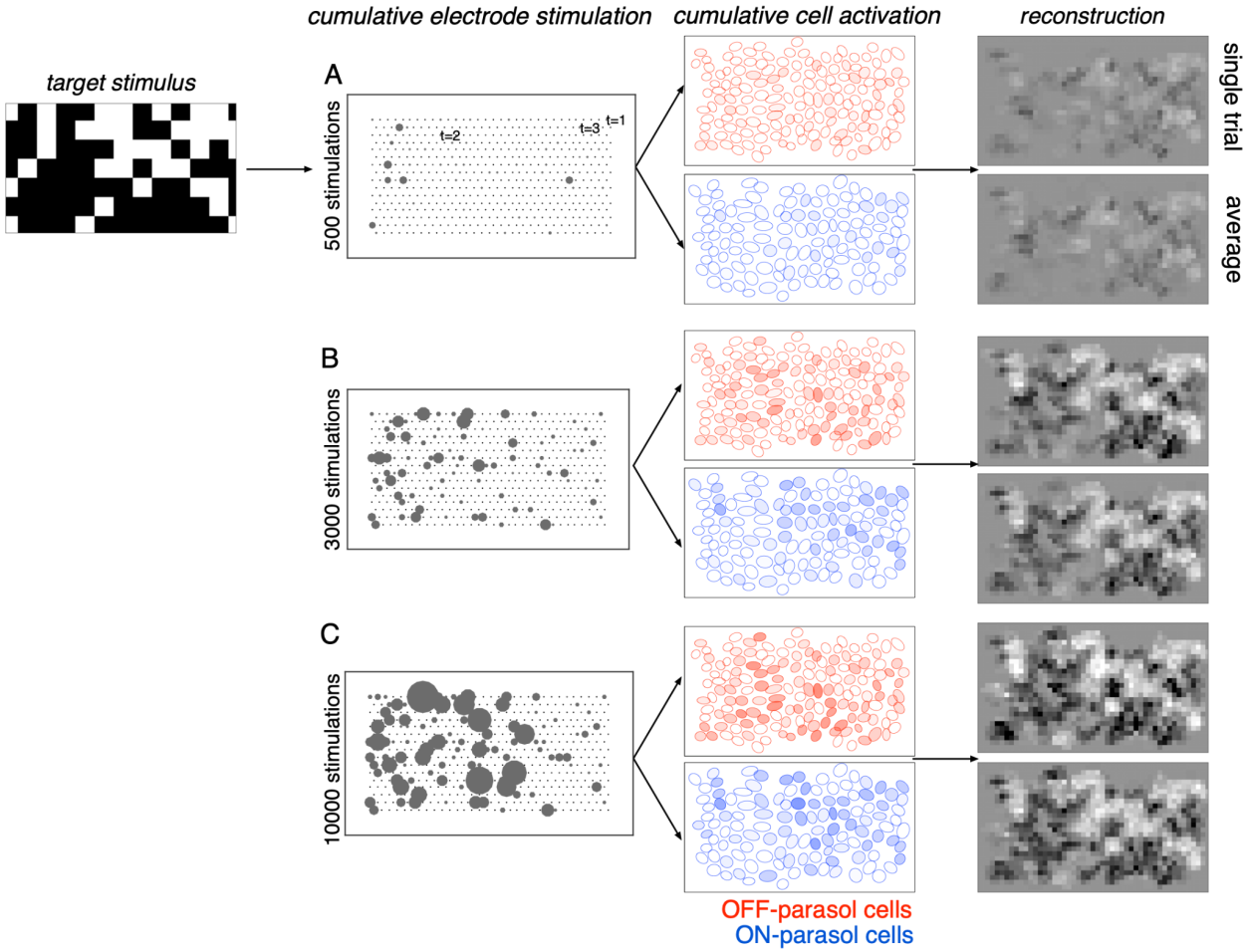

Figure 2: **Stimulus reconstruction achieved using the greedy dithering algorithm:** White noise target image shown on left. First column: Cumulative stimulation count across electrodes after 500, 3000 and 10000 electrical stimuli (A, B, C respectively). Second column: Responses for ON (blue) and OFF (red) parasol cells, sampled according to the response probability associated with the electrical stimuli used. Shade indicates the cumulative number of spikes. Third column: Single-trial and trial-averaged reconstruction of the target stimulus.

Greedy temporal dithering is nearly optimal given the interface constraints

Could the performance of the approach be improved? Performance could in principle be limited by the greedy selection of electrical stimuli or by the limited control of neural activity afforded by single-electrode stimulation.

To test whether performance could be improved with a different algorithm, the greedy approach was compared with a nearly ideal algorithm. The original optimization problem can be reformulated as:

$$minimize_{w \geq 0} \; ||S - ADw||^2 + V^T w \; such\; that \; w \in Z_+$$

where S is the target stimulus, A is reconstruction filter, D is a matrix of all response probabilities in the dictionary and $w \geq 0$ indicates the number of times each dictionary element is used. Because w is an integer, the general optimization problem is difficult to solve. However, an upper bound on the performance gap between the greedy dithering algorithm and the optimal algorithm can be obtained by allowing non-integer values of w. Across multiple target images, the gap was low (<10%, "optimal algorithm" in figure 3E,H), suggesting that the approximate nature of the greedy algorithm is not a significant source of error in the present conditions.

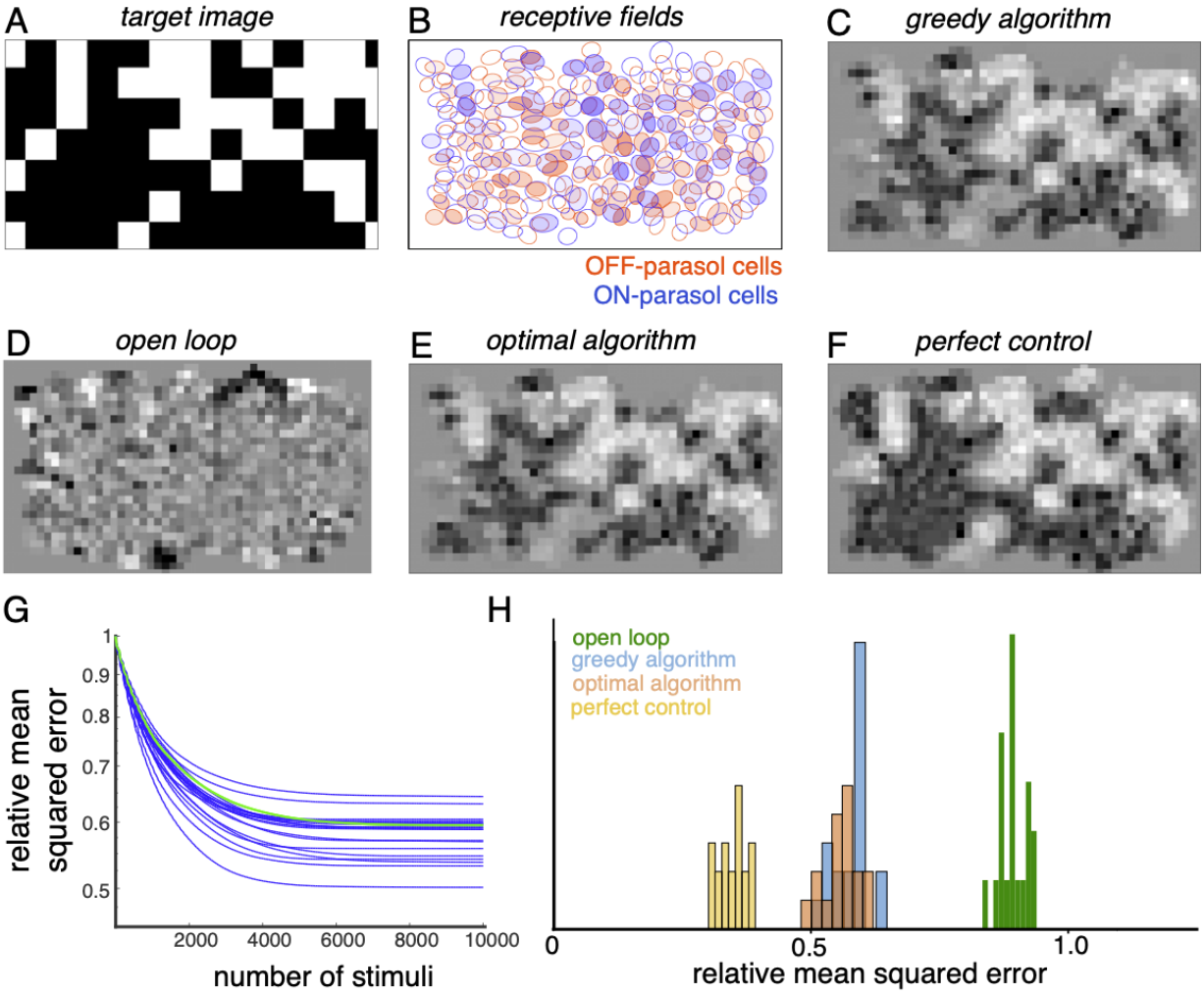

Figure 3: **Quantifying the performance of the proposed stimulation approach.** (A) A sample target white noise image. (B) ON and OFF receptive fields shaded with the expected summed response from greedy dithering. Achieved reconstructions are shown using greedy dithering (C), open-loop control (D), a

lower bound on the optimal algorithm for a single-electrode dictionary (E) and perfect control with the available reconstruction filters (F) . (G) Reconstruction error (relative MSE) between target and the achieved perception for 20 different targets (blue lines), with the example from A-F indicated with the green line. (H) Histogram of relative performance of the above approaches across 20 target images.

To test whether performance could be improved with a more precise neural interface, the reconstruction error was compared with ideal control, in which any desired response pattern can be produced. The performance with ideal control was estimated by the solving the following optimization problem:

$$minimize_{r \geq 0} ||S - Ar||^2 \; such \; that \; r \; \in Z_+$$

where r is the vector of spike counts. This optimization problem was solved by relaxing the integer constraint on r and evaluating the gap between the performance using a single-electrode dictionary and an ideal dictionary. Across multiple targets, this gap was substantial (more than 40%, “perfect control” in Figure 3F, H).

Thus, although the greedy algorithm is nearly optimal for a single-electrode dictionary, the reconstruction performance of an artificial retina could be improved by enhancing the dictionary (for example, by using calibrated multi-electrode stimulation patterns).

Spatial multiplexing for fitting multiple stimuli in a visual integration window

The main requirement of temporal dithering – the independence of responses generated by individual electrical stimuli, and their summation within a time window for visual perception – could limit the throughput of electrical stimulation. Although independence can be ensured by spacing single-electrode stimuli widely in time (e.g. several ms), this approach could make it difficult to deliver many electrical stimuli within a visual integration window (e.g. tens of ms). One approach to maximizing the number of stimuli that can be delivered within a given time window is *spatial multiplexing*, in which multiple electrodes that influence firing independently of one another are used simultaneously. For independence to hold, the following condition must be met: if p_1 is the activation probability of a given cell with stimulation on electrode 1, and p_2 is the activation probability of the same cell with electrode 2, then the activation probability with simultaneous stimulation must be $p_1 + p_2$. This condition was ensured using two simple approaches.

First, stimulation using two electrodes and current levels that produce $p_1 > 0$ and $p_2 > 0$ is not allowed. This is implemented in the algorithm by updating the electrical stimulation dictionary after each greedy choice to disallow the selection of electrodes that would produce nonzero activation probability of cells that were activated recently.

Second, if $p_1 > 0$ and $p_2 = 0$, simultaneous stimulation is allowed only if stimulation using both electrodes yields activation probability p_1. This condition was ensured by incorporating a spatial exclusion zone for simultaneous stimulation (Figure 4A). To estimate the spatial exclusion zone, activation curves corresponding to single-electrode stimulation near the soma were compared to activation curves obtained with additional simultaneous stimulation using a nearby secondary electrode (see Methods). Examination of many electrode pairs at varying distances (Figure 4B) demonstrated a systematic decrease in interaction between stimulating electrodes with distance. On average, the activation probability of the primary electrode was affected relatively little (<4% fractional change in threshold) if the secondary electrode was more than 300 μm away. Although this is not an exhaustive test of independence, it gives an empirical estimate of the distance beyond which two electrodes are unlikely to interact in activating a cell.

Greedy dithering was combined with spatial multiplexing as follows. Multiple dictionary elements were chosen greedily at each time step, based on the strategy outlined above, and at the subsequent time step, the dictionary was updated to exclude elements that stimulated cells activated recently (Figure 4A). This approach was tested in a closed-loop experiment, using calibrated electrical and visual response properties. The greedy stimulation sequence was determined with the constraint that simultaneously stimulated electrodes must be least 160 μm apart, which produced little interaction (<7% fractional change in threshold; Figure 4B). For the chosen stimulation pattern, the average minimum distance between simultaneously activated electrodes was ~193 μm, and the relationship between stimulation amplitudes and observed activation probabilities was largely preserved compared to the calibrated responses to single-electrode stimulation (Figure 4C). This suggests that spatial multiplexing of electrodes outside a spatial exclusion radius is a practical strategy for high-throughput stimulation with temporal dithering.

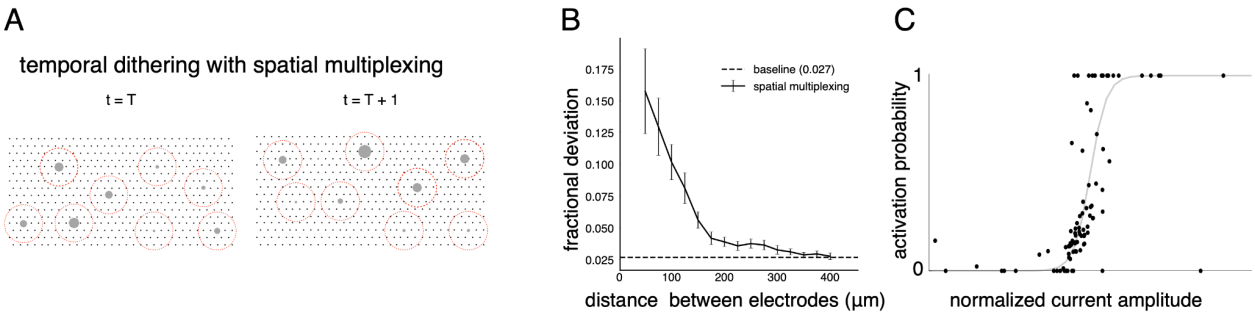

Figure 4: **Spatial multiplexing by simultaneous stimulation of distant electrodes.** (A) Visualization of temporally dithered and spatially multiplexed stimulation. At each time step, multiple single-electrode stimuli are delivered (gray circles) over the electrode array (black dots), separated by a spatial exclusion radius (red circles). (B) Estimation of the spatial exclusion radius. Interaction between electrodes is measured by fractional deviation in activation threshold for a given cell on a primary electrode (ordinate) resulting from simultaneous stimulation of another electrode with identical current amplitude at varying separations (abscissa). Mean fractional deviation and standard error for each distance was calculated using 754 cell-electrode pairs across 4 peripheral retinal preparations (see Methods). Baseline represents

the error of estimating single-electrode thresholds, computed using faraway secondary electrodes. (C) Closed-loop validation of spatial exclusion zone. Dots indicate empirical activation probability during greedy temporal dithering with spatial multiplexing. Solid curve indicates the activation curve measured during calibration of single-electrode responses, normalized across 52 cell-electrode pairs. Eight electrodes with at least 140 μm separation were stimulated every 2ms.

Greedy dithering framework enables data-driven hardware design

By framing the objective in terms of stimulus reconstruction, the greedy dithering framework suggests further optimizations for hardware efficiency while preserving performance. Because the greedy dithering algorithm chooses electrodes in a spatially non-uniform manner over the array (Figure 5A), restricting stimulation to a more frequently chosen subset of electrodes could provide this efficiency. To test this possibility, the algorithm was applied with various dictionaries, each restricted to a smaller subset of the most frequently used electrodes. Across 20 new targets, a minimal (<5%) increase in reconstruction error was observed if the number of available electrodes was reduced by less than 50% (Figure 5B-F). Note that this increase was not due to the greedy nature of the stimulation, because a lower bound computed for an optimal algorithm showed similar behavior (Figure 5B). These observations suggest a strategy for efficient implant operation in a retina-specific manner: identify the most frequently used ~50% of electrodes during calibration and permanently turn off the remaining electrodes during run-time usage. Such a reduction in the set of stimulated electrodes could lead to reduced memory access and power consumption. Thus, applying the algorithmic framework on the *ex vivo* lab prototype leads to insights relevant for the development of an *in vivo* implant.

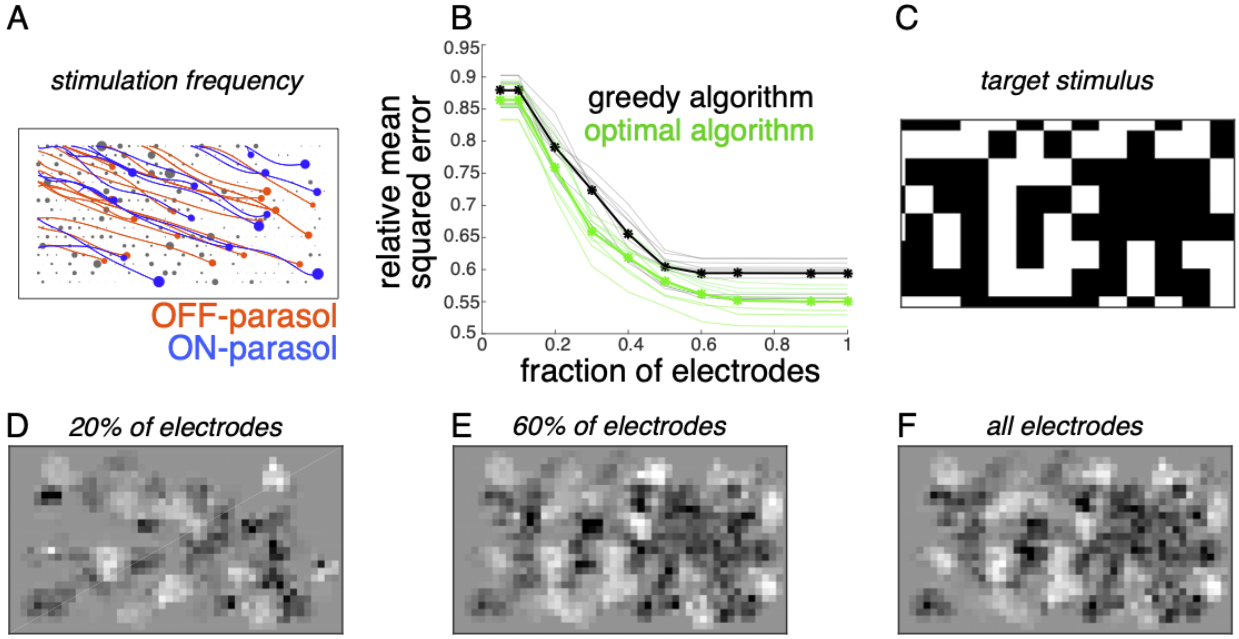

Figure 5: **Subsampling electrodes for hardware efficiency.** (A) Frequency of stimulating different electrodes (size of gray circles), overlaid with axons (lines) and somas (colored circles) inferred from spatiotemporal spike waveform across the electrode array recorded from each cell. (B) Reconstruction

error as a function of the fraction of electrodes included in the dictionary (black, thin lines correspond to different target images) and average over 20 target images (black, thick line). Different collections of target stimuli were used for electrode selection and reconstruction performance evaluation. Lower bound on error of any algorithm for the subsampled dictionaries for individual targets (green, thin lines) and averaged across targets (green, thick line). (C) Example target image. (D-F) Reconstructed images using the dictionary with most frequently used 20%, 60% and 100% of electrodes respectively.

Greedy dithering framework extends to naturalistic viewing conditions

For practical application, the stimulation algorithm must be extended to naturalistic viewing conditions, in which saccadic and fixational eye movements normally move the high-resolution fovea over the scene (Figure 6A). Similarly, the implant will presumably move over the scene as the eye moves, and should only transmit the information about its restricted view of the image. The greedy dithering framework extends naturally to this condition.

First, the objective, the constraint and an algorithm for the corresponding optimization problem are identified. The objective function is modified to minimize the error between the original stimulus movie and a movie reconstructed from the RGC spike trains. For simplicity, a spatio-temporal reconstruction filter is used with separable spatial and temporal components and the same time course for ON and OFF parasol cells (Figure 6B, Methods). The constraints (measured electrical stimulation properties) are unchanged. Finally, the greedy algorithm is adapted by choosing a dictionary element for each time step to minimize the average error between the recent frames of the target stimulus and the corresponding frames of the reconstruction (Methods).

This modified algorithm was evaluated using simulations of naturalistic viewing. For a given visual scene, a dynamic visual stimulus was generated by simulating saccadic eye movements with random inter-saccade intervals and random fixation locations with a preference for regions of the scene with high spatial frequency content (Yarbus 1967). Optionally, fixational eye movements were simulated by jittering the visual stimulus with brownian motion (see Methods). As before, stimulation patterns were determined using the greedy dithering algorithm, and evoked responses were simulated using the measured single-electrode activation probabilities. The dynamic stimulus over the array was reconstructed from RGC spikes and the visual scene was then assembled after re-centering each frame of the reconstruction relative to the array.

The assembled visual scene closely matched the target, capturing many of the fine details not captured in existing open loop stimulation (Figure 6C, Supplementary movie 1). Interestingly, the reconstructed visual scene was smoother and more accurate (lower reconstruction error) when fixational eye movements were simulated along with saccades (Figure 6C). Specifically, for the same final reconstruction error, a ~4X reduction in the number of required saccades and hence the number of required electrical stimuli was observed in the presence of fixational eye movements (Figure 6D). Hence, the modified greedy dithering approach can improve

reconstruction performance in natural viewing conditions and reveals more accurate image reconstruction with fixational eye movements.

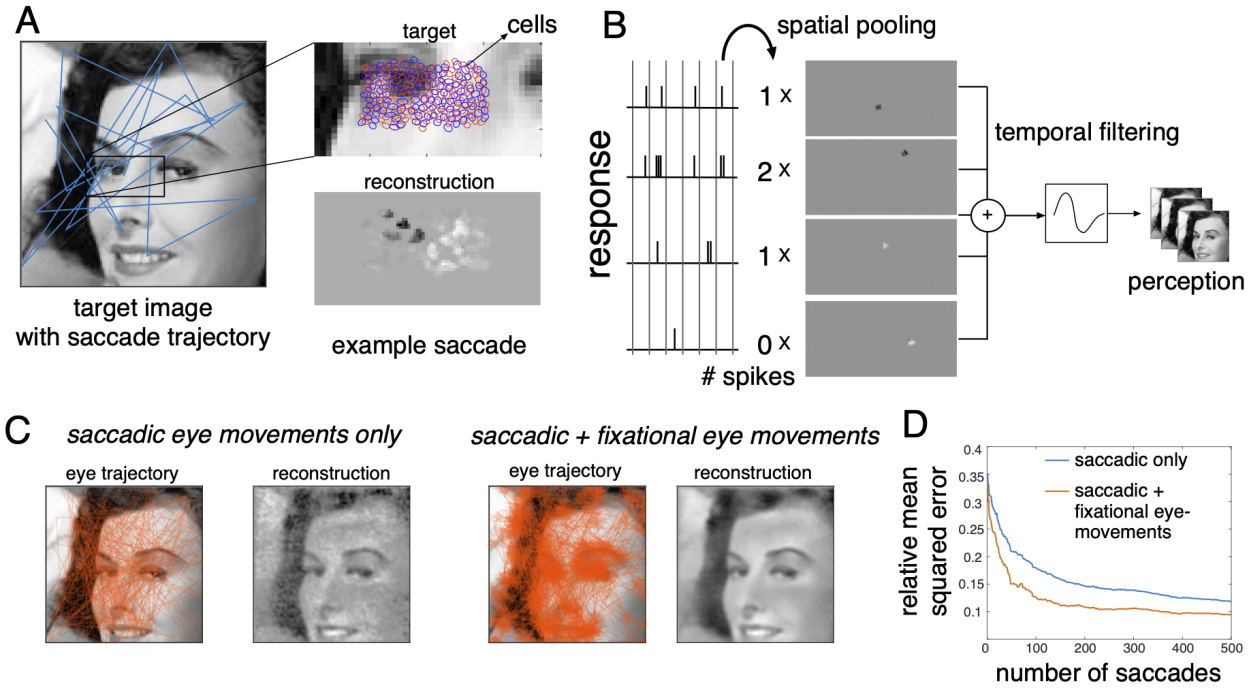

Figure 6: **Extension of the greedy algorithm to naturalistic conditions with eye movements.** (A) Conversion of a visual scene into dynamic stimulus. A target visual scene (left), with sample eye movement trajectory (blue). For each eye position, the population of ganglion cells accessible by the implant views a small portion of the visual scene (top right). The reconstructed stimulus for each patch captures the local stimulus information (bottom right). (B) Spike trains passed through a spatio-temporal reconstruction filter of the dynamic stimulus movie. For simplicity, a rank one filter was used, which spatially filtered each spike bin independently, and then filtered the reconstructed stimulus movie in time. (C) Final reconstruction performance over a sequence of saccades, in the absence (left) and the presence (right) of small fixational eye movements. (D) Reduction in reconstruction error of the visual scene as a function of the number of saccades ,in the absence (blue) and the presence (orange) of fixational eye-movements.

Optimizing stimulation using a perceptual similarity measure

The framework provides a way to use alternative metrics to optimize visual perception evoked by electrical stimulation. Specifically, the mean square error measure of reconstruction accuracy, while convenient, does not accurately capture perceived differences in image content, whereas metrics such as Structural Similarity (SSIM) more closely parallel perception (Wang et al. 2004). To identify a nearly optimal sequence of stimuli with SSIM as the objective, an exhaustive approach was used to optimize across all possible stimulation sequences for every eye location (see Methods). The SSIM and MSE metrics produced similar reconstructions when the number of electrical stimulation patterns was unlimited (Figure 7B). This suggests that the choice of reconstruction error metric may not be important for an implant that can stimulate at

high rates. However, SSIM optimization produced higher quality reconstructions when the number of electrical stimuli was limited (Figure 7C), a constraint that could apply in real-world usage because of heat dissipation limits in an implant. Thus, the greedy dithering approach with a perceptually accurate reconstruction metric could lead to better performance in an implanted device, though additional developments will be needed before such an optimization can be performed in real time.

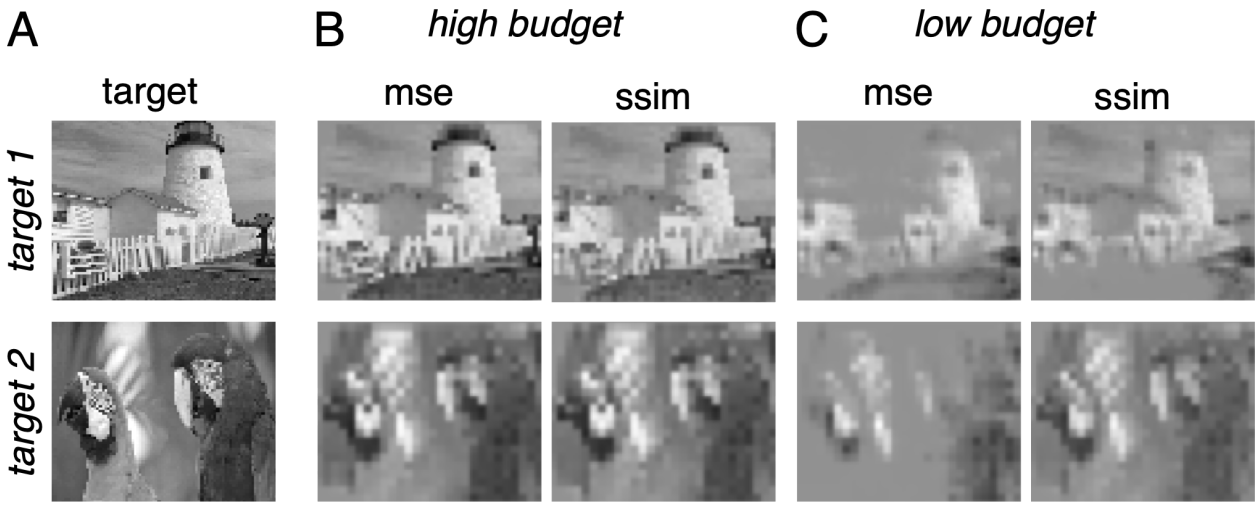

Figure 7: **Extension of the stimulation algorithm using Structural Similarity (SSIM) perceptual error metric.** (A) Two target images. (B) with the reconstruction with MSE and SSIM error measures for greedy dithering, with a (high) budget. (C) Same as B, with a low budget.

Discussion

This paper presents a dynamic algorithmic approach to improve the performance of sensory electronic implants. Greedy dithering and multiplexing overcomes the challenges of precisely controlling the activity of neural populations using electrical stimulation by rapidly delivering a sequence of electrical stimuli with independent effects within a visual integration window. This approach avoids non-linear interactions resulting from simultaneous multi-electrode stimulation while providing enough flexibility to elicit rich spatio-temporal responses using single-electrode stimulation. The greedy dithering and multiplexing approach outperforms existing open loop approaches (Figure 2, 3, 4) and enables efficient neural interface development using a small collection of calibrated stimulation patterns (Figure 5), potentially incorporating naturalistic viewing with eye movements (Figure 6) and/or perceptual similarity metrics (Figure 7).

The presented approach relies on several assumptions regarding how the brain uses RGC responses for vision. A major assumption of the approach is that downstream processing of RGC responses is slow, with a time constant of tens to hundreds of ms depending on the task, so that perception only depends on the total number of spikes within this time interval (Wandell 1995). Another assumption is that the brain interprets the spike counts by explicitly

reconstructing the visual stimulus. For simplicity, linear reconstruction was assumed, with a filter that minimizes mean-squared error. A more accurate approach could involve replacing this with a non-linear and/or biologically realistic reconstruction (Parthasarathy et al. 2017; Kim et al. 2021; Wu et al. 2022) potentially using a more perceptually accurate stimulus similarity metric (Figure 7) (Wang et al. 2004; Shah and Chichilnisky 2020).

The modular nature of the algorithmic approach enables several potential improvements in parallel. The single-electrode dictionary could be enhanced with electrical stimulation patterns designed to optimize cellular selectivity, response diversity, or ideally the overall expected algorithm performance (Jepson et al. 2014; Fan et al. 2019; Vilkhu et al. 2021). Additionally, the efficient and real-time greedy algorithm could be replaced with an algorithm that identifies the optimal stimulation sequence for multiple time-steps, perhaps accounting for predicted future saccade locations. Finally, each module could be optimized for metrics such as performance, hardware efficiency (e.g. Figure 5B) or stability/robustness for chronic function.

Implications for design of neural interfaces

The greedy dithering approach relies on the ability to efficiently compute and deliver optimal electrical stimuli in real time in a bi-directional electronic retinal implant. Although this procedure exceeds the capabilities of current devices, two features of the approach could potentially make implementation possible on limited hardware.

First, the hardware requirements of the proposed closed-loop procedure are less stringent than those of a real-time system. In the latter, one records and analyzes the results of every electrical stimulus to determine the stimulus for the next time step. Instead, the present approach relies only on identifying the *average* electrical response properties of each cell in advance. After this initial calibration, the stimulation sequence is decided in an open loop manner by optimizing the expected visual reconstruction associated with each stimulus.

Second, the greedy dithering approach can effectively exploit non-selective activation. Selectively activation of every cell in a region of the retina, if achievable, would make it possible to create arbitrary patterns of neural activity, but in practice this is difficult with real neural interfaces. The presented framework uses all available stimulation patterns as efficiently as possible by directly optimizing for stimulus reconstruction. In fact, the approach frequently exploits non-selective activation, in order to evoke desired spiking activity in fewer time steps (Figure 5).

Translational potential

The physiological similarities between human and macaque retina (Cowan et al. 2019; Kling et al. 2020; Soto et al. 2020; Rodieck 1998), including their responses to electrical stimulation (Madugula et al. 2020), suggest that the benefits of the present stimulation approach could

translate from the *ex vivo* lab prototype with healthy macaque retina to an *in vivo* implant in the degenerate human retina. However, several technical innovations are required to enable chronic *in vivo* recording and stimulation in the degenerate retina. First, new surgical methods must be developed to implant a tiny chip on the surface of the retina with stable contact. Second, receptive field locations, cell types and reconstruction filters must be inferred using electrical features rather than light-evoked responses in a blind retina (Sekirnjak et al. 2011; Li et al. 2015; Richard, Goetz, and Chichilnisky 2015; Zaidi et al. 2022). Third, the stimulation approach must be modified to account for changes in spontaneous/oscillatory activity in the degenerated retina (Sekirnjak et al. 2011; Goo et al. 2015; Trenholm and Awatramani 2015). Finally, the approach must be tested with the visual and electrical properties in the central retina (Gogliettino et al. 2022), the most clinically relevant location for a retinal implant.

The present approach could also leverage other methods developed for improving the performance of existing, low resolution implants. Examples of these methods are context-dependent image preprocessing (Cha, Horch, and Normann 1992; McCarthy, Barnes, and Lieby 2011; Lieby et al. 2011; Vergnieux, Macé, and Jouffrais 2017; Ho, Boffa, and Palanker 2019), limiting to sparse stimulation (Loudin et al. 2007), exploiting the adaptation of sensory systems (Rouger et al. 2007; Merabet and Pascual-Leone 2010) and exploiting perceived phosphenes due to axon bundle activation for optimizing stimulation (Granley, Relic, and Beyeler 2022; de Ruyter van Steveninck et al. 2022; Relic et al. 2022). Ideally, a unified framework such as the one presented here would include these and potentially other approaches to optimal stimulation.

Outside the retina, recent work on electrical stimulation of the human visual cortex (Beauchamp et al. 2020) applied a dynamic approach in a visual cortical implant and demonstrated impressive performance in human participants. However this work only considered simple visual stimuli which can be described by lines (such as English letters and numbers). The integrated and modular framework presented here could provide a way to precisely control neural activity for arbitrarily complex stimuli and improve the performance of cortical sensory implants.

Methods

Retinal preparation.

Extracellular multi-electrode recording and stimulation in macaque retina were performed as described previously (Jepson et al. 2013, Grossberg et al., 2018). Briefly, eyes were obtained from terminally anesthetized macaque monkeys used for experiments in other laboratories, in accordance with Institutional Animal Care and Use Committee guidelines. After enucleation, the eyes were hemisected and the vitreous humor was removed. The hemisected eye cups containing the retinas were stored in oxygenated bicarbonate-buffered Ames' solution (Sigma) during transport to the laboratory. The retina was then isolated from the pigment epithelium under infrared illumination and held RGC-side down on a custom multielectrode array (see below). Throughout the experiments, the retina was superfused with Ames' solution at 35°C.

Recordings

A custom 512-electrode stimulation and recording system (Hottowy et al. 2008, 2012) was used to deliver electrical stimuli and record spikes from RGCs. The electrodes were organized in a 16×32 grid with isosceles triangular lattice arrangement, with 60 μm spacing between electrodes (Litke et al. 2004). Electrodes were 10 μm in diameter and electroplated with platinum. For recording, raw voltage signals from the electrodes were amplified, filtered (43–5,000 Hz), and multiplexed with custom circuitry. These voltage signals were sampled with commercial data acquisition hardware (National Instruments) at 20 kHz per channel.

Estimation of visual stimulus reconstruction filter

Recordings obtained with visual stimulation were analyzed to identify spike waveforms of distinct RGCs in the absence of electrical stimulation artifact, using spike sorting methods described previously (Field et al. 2007; Litke et al. 2004), which identified spike times of identified RGCs on the basis of relatively large, stereotyped spikes detected near the soma. The complete spatiotemporal signature of the spikes from each cell over all electrodes (the *electrical image*) was then computed by averaging the voltage waveforms on all electrodes at and near the times of its recorded spikes (Litke et al. 2004). The electrical image of each cell provided a template of its spike waveform. This was used to identify the cells producing spikes in response to electrical stimulation.

Distinct RGC types were identified by their distinct responses to a 30 minute long white noise visual stimuli (80x40 pixel grid, ~42 μm pixels, refresh rate 120 Hz). Briefly, a dynamic random checkerboard stimulus was presented, and the average stimulus that preceded a spike in each RGC was computed, producing the spike-triggered average (STA) stimulus (Chichilnisky 2001). The STA summarizes the spatial, temporal, and chromatic properties of light responses. Features of the STA were used to segregate functionally distinct RGC classes. Spatial receptive fields for each cell type were obtained from fits to the STA (Chichilnisky and Kalmar 2002). For each identified RGC type, the receptive fields formed a regular mosaic covering the region of retina recorded (Devries and Baylor 1997; Field et al. 2007), confirming the correspondence to a morphologically distinct RGC type (Dacey 1993; Wässle et al. 1981), and in some cases revealing complete recordings from the population. The density and light responses of the five most frequently recorded RGC types uniquely identified them as ON and OFF midget, ON and OFF parasol and small bistratified Cells (SBCs). Subsequent analysis was restricted to two numerically dominant RGC types – ON parasol and OFF parasol cells – which are sampled efficiently in our experiment and formed nearly complete mosaics covering the region recorded.

For each cell, the stimulus reconstruction filter was approximated using the scaled receptive field. The receptive field was obtained by computing the spatial component of the rank 1 approximation of the STA. The receptive field was then denoised by computing the robust standard deviation (σ) of magnitudes of all pixels, zeroing out pixels with absolute value less than 2.5σ, and retaining the largest spatially contiguous component. The receptive field was scaled such that a linear-rectified model of responses to static flashes of checkerboard matched the observed average number of spikes across recordings. The scale of the receptive field for the stimulus reconstruction filter was estimated for optimally reconstructing the static stimuli from responses of the linear-rectified model.

Note that the cell types and reconstruction filters were identified from light-evoked responses for convenience. In a clinical implant, this would not be possible. Previous work has shown that distinct cell types can be identified from spontaneous electrical activity (Richard, Goetz, and Chichilnisky 2015; Zaidi et al. 2022)

Estimation of electrical stimulus dictionary

For electrical stimulation, custom hardware (Hottowy et al. 2012) was controlled by commercial multifunction cards (National Instruments). Current was passed through each of the 512 electrodes individually, with 40 different amplitudes (0.1 μA - 4 μA), 27 times each. For each amplitude, charge-balanced tri-phasic current pulses with relative amplitudes of 2:−3:1 and phase widths of 50 μs (total duration 150 μs) were delivered through the stimulating electrode (amplitude corresponds to the magnitude of the second, cathodal phase of the pulse). This pulse shape was chosen to reduce stimulation artifacts in the recordings. Custom circuitry disconnected the recording amplifiers during stimulation, reducing stimulation artifacts and

making it possible to identify elicited spikes on the stimulating electrode as well as nearby electrodes (Hottowy et al. 2012; Jepson et al. 2013). For recording and stimulation, a platinum ground wire circling the recording chamber served as a distant ground.

Response probabilities for each stimulation pattern were identified after removing electrical artifacts using custom spike sorting software (Mena at al., 2017). Briefly, the spike sorting software estimates the electrical stimulation artifacts by modeling the artifact change across amplitudes with a Gaussian process, subtracts the estimated electrical artifacts and then matches spikes in residual recordings to cell waveforms estimated from visual stimulus recordings. The cell activation probabilities for each of the 40 different amplitudes were smoothed with a sigmoid fit and collected in a dictionary. Each item in the dictionary consisted of an electrode, a stimulus current level, and the evoked spike probability of all recorded cells (most cells had an evoked spike probability of 0).

Dictionary elements that involved activating cells along their axons, with somas off the electrode array, were removed due to their unknown receptive field locations and thus uncertain contribution to stimulus reconstruction (Grosberg et al., 2017). Briefly, the responses to electrical stimulation were mapped to a collection of weighted graphs, and graph partitioning and graph traversal algorithms were applied to identify bundle activity. The focus was on two characteristic features of axon bundle signals: bidirectional propagation, and growth of signal amplitude with stimulation current (Tandon et al. 2021) .

Finally, only dictionary elements that activated at least one cell with probability at least 0.01 were retained, resulting in 1000-5000 dictionary elements. A single dictionary element that does not activate any cell (probability = 0) was added to allow the greedy algorithm to avoid stimulation when any real stimulation pattern would increase error.

Greedy dithering algorithm

Given the stimulus reconstruction filter and electrical stimulation dictionary, the goal of the greedy dithering algorithm is to identify a sequence of electrical stimuli that encodes a target visual stimulus.

Let S be the target visual stimulus, A be the stimulus reconstruction filter and R_i be the observed response vector (consisting of a zero or a one for each cells) produced in the population of cells stimulated using dictionary element C_i with the associated probability vector p_i.

Multiple dictionary elements must be combined to generate rich spatio-temporal population responses corresponding to a target visual stimulus. Hence, the overall problem reduces to finding a sequence of dictionary elements $\{C_1, \cdots, C_T\}$ that minimizes the expected mean-squared error between the target visual stimulus and reconstructed responses:

$$C_1, \ldots, C_T = argmin \; E_{R_i \sim Bernoulli(p_{c_i})} \| S - A(R_1 + \ldots + R_T) \|^2$$

Instead of jointly optimizing for the whole stimulation sequence, which is difficult, a greedy approach is proposed in this paper. Using the independence of responses across electrical stimuli, the mean-squared error term is decomposed into two terms corresponding to bias and variance. The variance term ($tr(\, var(AR_i))$) corresponds to the total variance across all pixels of the visual image for the stimulation pattern C_i.

$$C_1, \ldots, C_T = argmin \; \| S - A(p_1 + \ldots + p_T) \|^2 + tr(\, var(AR_i)) + \cdots + tr(var(AR_T))$$

The greedy choice at step t only depends on the response probabilities associated with previous stimulation up to step $t - 1$, and the contribution to the variance from previous stimulation choices is constant for all possible dictionary elements. Hence, the greedy objective function for choosing the stimulation pattern at step t is given by:

$$C_t = argmin_c \; \| S - A(p_1 + \ldots + p_{t-1} + p_c) \|^2 + tr(var(AR_c))$$

The variance term can be computed explicitly for the Bernoulli distribution of evoked responses.

The above expressions can be modified to incorporate biological and hardware constraints that change during the course of the stimulation sequence. The modified greedy choice is given by the following equation, where D_t denotes the choice of dictionary elements available at time t, based on stimulation patterns chosen for previous time steps ($C_1, \cdots, C_{t-1}$).

$$C_t = argmin_{c \in D_t} \; \| S - A(p_1 + \ldots + p_{t-1} + p_c) \|^2 + tr(var(AR_c))$$

Characterizing spatial exclusion radius for spatial multiplexing

The spatial exclusion radius was estimated using a bi-electrode stimulation experiment as follows. The initial response dictionary was characterized using single-electrode stimulation as described above. A target cell was chosen, and the activation curve was determined for the electrode that recorded the largest amplitude spike waveform (primary stimulating electrode). Equal current was then passed simultaneously through a secondary electrode, and the changes in the activation curve were examined. The secondary electrode was varied over all electrodes within 400 μm of the primary electrode. As a control, a set of 15 secondary electrodes, each more than 800 μm from the primary electrode and not overlapping the axon of the target cell, were chosen and used to estimate the single-electrode activation curve in the same stimulation conditions.

For each bi-electrode stimulation, identical currents were passed through the primary and secondary electrodes, at 40 different log-spaced current amplitudes from 0.1-4 μA, 25 times

each. This fixed-ratio stimulation design choice limited the scan time to a practical duration (2500 sec) and provided a conservative estimate of the potential impact of a secondary electrode on activation by the first electrode.

Activity was recorded following every stimulation. Evoked spikes were identified and separated from stimulation artifacts as follows. First, for each stimulation pattern and amplitude, all recorded traces were collected over the primary electrode and surrounding electrodes. Second, raw recordings were clustered into distinct events. Third, difference signals between events were matched automatically to the previously identified candidate spike waveforms (Madugula et al. 2020). This approach exploits the regularity of artifacts across traces for each stimulation pattern and amplitude pair.

The two-electrode stimulation produced an activation curve for each electrode pair, from which the activation threshold (current amplitude for 50% spike probability) was determined. Since the saturation of some recording electrodes could produce an error in the activation curve estimates, the analysis was limited to current patterns that showed stable activation curves when spike sorting was performed with and without the saturated electrodes. Any such electrode pair with inconsistent activation thresholds was excluded from further analysis. The mean activation threshold across all control stimulation patterns defined the single-electrode activation threshold to which each of the test stimulation patterns was compared. In particular, the fractional change from the single-electrode activation threshold was computed for each test stimulation pattern. This deviation indicates the degree to which the presence of a secondary stimulating electrode influences the responses generated by a particular primary electrode.

The experiment described above was performed in 4 peripheral macaque retinal preparations, targeting 5 ON and 2 OFF parasol cells, for a total of 754 test pairs of primary and secondary electrodes. Figure 4B summarizes the absolute change in threshold with increasing distance between stimulating electrodes, generated by computing the weighted mean for test pairs close to each distance. The weighting for each test pair was inversely proportional to the variance of activation thresholds from the corresponding control set of distant primary and secondary electrodes. This weighting scheme accounts for variability in electrical activation properties across cells and retinal preparations. The resampled standard error of the weighted mean was computed at each distance. The baseline measure is computed by repeating this procedure for all control electrode pairs.

Closed loop validation of temporal dithering & spatial multiplexing

The spatio-temporal dithering approach was tested using a closed loop experiment in the lab prototype as follows. Initial calibration was performed using a white noise visual stimulus (30 min) and an exhaustive single-electrode stimulation scan (4400 sec, 27 trials per electrode and amplitude combination, 40 amplitudes at each electrode logarithmically in the range 0.1-4 μA). After calibration, a greedily chosen stimulation sequence was delivered, with spatial and

temporal multiplexing to optimize image reconstruction for different target images. After greedy stimulation, a white noise visual stimulus and electrical stimulation scans were repeated to evaluate the stability of the retina during the course of the experiment. Finally, to aid in spike sorting, the greedy stimulation sequence was repeated after administering TTX to the preparation, to record and later subtract electrical artifacts (see below).

The cell locations, types and spike waveforms were identified from the recording with the visual stimulus using standard approaches. For simplicity, the linear reconstruction filter for each cell was assumed to be proportional to the spatial receptive fields estimated using the spike triggered average (Chichilnisky 2001). The spike waveforms were then used to estimate the electrically-evoked activation probability for each cell (Mena et al., 2017) and axon bundle activation threshold (Tandon et al. 2021). While the calibration runs were analyzed, static white noise images were presented to the retina (500 ms presentation, 250 ms gray screen at the same average intensity between flashes). The average spike count in response to these flashes was used to identify the scale of a linear reconstruction filter. Note that only ON and OFF parasol cells (180 cells) were considered for building the reconstruction filter and the dictionary.

The sequence of 4000 elements was identified for each of a collection of 20 random checkerboard stimuli using the estimated linear reconstruction filter and stimulation dictionary. The stimulation sequence was chosen to greedily minimize the expected mean-squared error at each step with additional constraints for avoiding temporal and spatial interactions (see Results).

The temporal constraint corresponded to avoid stimulating a cell again during its refractory period. After selecting a dictionary element, the set of potentially activated cells is identified (activation probability > 0.1), and any dictionary element that activated any of these cells with probability > 0.1 was removed for the next 100 stimuli. The spatial constraint avoids simultaneous activation of nearby electrodes to avoid nonlinear interactions. After selecting a dictionary element at a given time step, any other dictionary element with electrodes with 140 micrometers distance are removed for 8 subsequent steps. Spatial multiplexing was tested by simultaneous stimulation of eight elements every 2 ms. The stimulation sequence for each target was repeated 30 times.

Because recordings after electrical stimulation generally contain electrical artifacts which obscure spike waveforms, experimental estimation of these artifacts was performed. The preparation was administered with TTX (1mM, 60 µL dosage), and suppression of activity was verified by a decline in spontaneous activity as well as lack of light responses after 15 minutes. Subsequently, the three stimulation sequences were repeated, with 10 repeats rather than 30 repeats. The spike sorting analysis after subtracting TTX traces is described below.

Spike sorting temporally dithered and spatially multiplexed stimulation

Spike sorting entails identifying when each cell spikes in the spatio-temporally dithered stimulation described above. This is accomplished using spike waveforms obtained with visual stimulation, along with subtracting the artifact using traces recorded in TTX.Specifically, spike sorting was accomplished as follows for each electrode and amplitude pair. First, all recorded traces over the stimulation electrode and 6 surrounding electrodes were obtained. Second, an estimate of the artifact computed by averaging traces from the corresponding TTX run were subtracted from the data. Third, spectral clustering was used to identify spikes, with the number of clusters identified by a gap in the Eigen spectrum. Fourth, each cluster average was associated with an identified spike from visual stimulation using binary pursuit template matching, using as candidates all neurons with spike amplitude <-30 DAC units on the stimulating electrode and <-45 DAC units on the surrounding electrodes. In previous work, template matching of clustered waveforms has shown improved agreement with manual labeling compared to template matching each waveform in isolation (Madugula et al. 2020).

To verify spike sorting, the average activation probability was computed across all electrical stimuli for each cell-electrode pair. This probability was then compared with the activation probability computed using the single-electrode scan. Additionally, only current levels that did not evoke off-array cell activation (axon bundle activation) were analyzed further (Tandon et al. 2021). Figure 4C compared the estimated activation probability from spatio-temporally dithered stimulation, with stimulation current normalized to the single-electrode scan for each cell-electrode pair. Only current patterns with stimulation frequency between 1-99 percentile and cells with amplitude <-35 DAC units on the stimulating electrode were considered.

Extension of greedy dithering to natural scenes with eye movements

The greedy temporal dithering approach was extended to natural viewing by modifications to visual stimulus target generation and reconstruction. For a given natural image, a dynamic visual target was generated by simulating eye movements. Nearly 500 fixation locations were sampled, preferentially at the high spatial-frequency regions of the image, and with a mean duration of 300 ms (SD 100 ms) between saccades. A patch of size 40 x 80 was taken around each saccade location to generate the dynamic visual stimulus. The dynamic visual stimulus seen by the region of retina targeted was simulated by a sequence of flashes, where each flash corresponds to a segment of the image with random time switching time between flashes. In some cases, fixational eye movements were also simulated by perturbing the fixation location with a brownian motion (3 pixel SD).

Given the dynamic visual target, the greedy algorithm is modified such that the stimulation choice at each step considers multiple recent frames of the target. The dynamic target was discretized on the display at 120Hz, and 83 stimulation choices were made within each frame (corresponding to a stimulation every ~0.1 ms). To accommodate the dynamic stimulus, the spatial reconstruction filter was replaced with a spatio-temporal reconstruction

filter. For efficiency, the spatio-temporal reconstruction filter was modeled as rank 1 (space-time separable), with the identical time course for all cells. Hence, each evoked spike influences the reconstruction at multiple subsequent time steps. The straightforward extension of the greedy algorithm is then to choose a stimulation pattern at each time step such that it minimizes the total error over multiple time steps.

For a given stimulation sequence, the image is assembled by first reconstructing each frame of the dynamic visual stimulus using the spatio-temporal filter. Then, each frame of the reconstructed dynamic stimulus is 'pasted' at the fixation location at the time of the spike. The intensity for each pixel in the final reconstructed image is estimated by averaging the intensity across all fixation locations in which the recorded cells have reconstruction filters that include the pixel.

Incorporating perceptual similarity metrics

Possible improvements to the approach that could be produced by optimizing perceptual similarity (rather than mean square error) in the stimulation objective were analyzed after simplifying modifications. First, instead of image-dependent and random fixation locations, all possible saccade locations were considered. This corresponds to a uniform distribution of fixation locations, and the visual scene is reconstructed by averaging the reconstruction of image patches corresponding to various fixation locations. Next, for each fixation location, the corresponding image patch was reconstructed using *expected* responses (rather than measured, stochastic responses). Note that unlike the algorithm presented above, this formulation does not account for inter-trial variability. Finally, instead of greedily optimizing the stimulation sequence, the number of stimuli for all dictionary elements and fixation locations were jointly optimized ($\{q_i\}$).

Given these simplifications, the following optimization problem is solved:

$$min_{\{q_i\}>0}\, d(S,\ G(\{ADq_i\}_{i=1}^{i=\#patches}))\ +\ \sum_i \lambda |q_i|_1$$

Where d is the measure of similarity, S is the target visual stimulus S and A is the reconstruction filter, D is the dictionary, q_i is the stimulation pattern of patch i, and G is an operator that averages the reconstruction of individual patches to assemble the entire image. To explore the reconstruction under different stimulation budgets, λ is varied to penalize stimulating a large number of dictionary elements.

Acknowledgements

We thank J. Carmena, K. Bankiewicz, T. Moore, W. Newsome, M. Taffe, T. Albright, E. Callaway, H. Fox, R. Krauzlis, S. Moriarty, and the California National Primate Research Center for access to macaque retinas. We thank the Stanford Artificial Retina team for helpful discussions. We thank Research to Prevent Blindness Stein Innovation Award, Wu Tsai Neurosciences Institute Big Ideas, NIH NEI

R01-EY021271, NIH NEI R01-EY029247, NIH NEI P30-EY019005 and NSF/CRCNS grant (EJC) for funding this work.

References

Armenta Salas, Michelle, Luke Bashford, Spencer Kellis, Matiar Jafari, Hyeongchan Jo, Daniel Kramer, Kathleen Shanfield, et al. 2018. “Proprioceptive and Cutaneous Sensations in Humans Elicited by Intracortical Microstimulation.” *eLife* 7 (April). https://doi.org/10.7554/eLife.32904.

Beauchamp, Michael S., Denise Oswalt, Ping Sun, Brett L. Foster, John F. Magnotti, Soroush Niketeghad, Nader Pouratian, William H. Bosking, and Daniel Yoshor. 2020. “Dynamic Stimulation of Visual Cortex Produces Form Vision in Sighted and Blind Humans.” *Cell* 181 (4): 774–83.e5.

Beyeler, Michael, Devyani Nanduri, James D. Weiland, Ariel Rokem, Geoffrey M. Boynton, and Ione Fine. 2019. “A Model of Ganglion Axon Pathways Accounts for Percepts Elicited by Retinal Implants.” *Scientific Reports* 9 (1): 9199.

Bloch, Edward, and Lyndon da Cruz. 2020. “The Argus II Retinal Prosthesis System.” *Prosthesis*. https://doi.org/10.5772/intechopen.84947.

Brackbill, Nora, Colleen Rhoades, Alexandra Kling, Nishal P. Shah, Alexander Sher, Alan M. Litke, and E. J. Chichilnisky. 2020. “Reconstruction of Natural Images from Responses of Primate Retinal Ganglion Cells.” *eLife* 9 (November). https://doi.org/10.7554/eLife.58516.

Cha, K., K. W. Horch, and R. A. Normann. 1992. “Mobility Performance with a Pixelized Vision System.” *Vision Research* 32 (7): 1367–72.

Chichilnisky, E. J. 2001. “A Simple White Noise Analysis of Neuronal Light Responses.” *Network* 12 (2): 199–213.

Cowan, Cameron S., Magdalena Renner, Brigitte Gross-Scherf, David Goldblum, Martin Munz, Jacek Krol, Tamas Szikra, et al. 2019. “Cell Types of the Human Retina and Its Organoids at Single-Cell Resolution: Developmental Convergence, Transcriptomic Identity, and Disease Map.” *bioRxiv*. https://doi.org/10.1101/703348.

Fan, Victoria H., Lauren E. Grosberg, Sasidhar S. Madugula, Pawel Hottowy, Wladyslaw Dabrowski, Alexander Sher, Alan M. Litke, and E. J. Chichilnisky. 2019. “Epiretinal Stimulation with Local Returns Enhances Selectivity at Cellular Resolution.” *Journal of Neural Engineering* 16 (2): 025001.

Flesher, Sharlene N., Jennifer L. Collinger, Stephen T. Foldes, Jeffrey M. Weiss, John E. Downey, Elizabeth C. Tyler-Kabara, Sliman J. Bensmaia, Andrew B. Schwartz, Michael L. Boninger, and Robert A. Gaunt. 2016. “Intracortical Microstimulation of Human Somatosensory Cortex.” *Science Translational Medicine* 8 (361): 361ra141.

Gaylor, James M., Gowri Raman, Mei Chung, Jounghee Lee, Madhumathi Rao, Joseph Lau, and Dennis S. Poe. 2013. “Cochlear Implantation in Adults: A Systematic Review and Meta-Analysis.” *JAMA Otolaryngology-- Head & Neck Surgery* 139 (3): 265–72.

Gogliettino, Alex R., Sasidhar S. Madugula, Lauren E. Grosberg, Ramandeep S. Vilkhu, Jeff Brown, Huy Nguyen, Alexandra Kling, et al. 2022. “High-Fidelity Restoration of Visual Signals by Electrical Stimulation in the Central Primate Retina.” *bioRxiv*. https://doi.org/10.1101/2022.05.24.493162.

Goo, Yong Sook, Dae Jin Park, Jung Ryul Ahn, and Solomon S. Senok. 2015. “Spontaneous Oscillatory Rhythms in the Degenerating Mouse Retina Modulate Retinal Ganglion Cell

Responses to Electrical Stimulation." *Frontiers in Cellular Neuroscience* 9: 512.

Granley, Jacob, Lucas Relic, and Michael Beyeler. 2022. "A Hybrid Neural Autoencoder for Sensory Neuroprostheses and Its Applications in Bionic Vision." *arXiv [cs.LG]*. arXiv. http://arxiv.org/abs/2205.13623.

Grosberg, Lauren E., Karthik Ganesan, Georges A. Goetz, Sasidhar S. Madugula, Nandita Bhaskhar, Victoria Fan, Peter Li, et al. 2017. "Activation of Ganglion Cells and Axon Bundles Using Epiretinal Electrical Stimulation." *Journal of Neurophysiology* 118 (3): 1457–71.

Ho, Elton, Jack Boffa, and Daniel Palanker. 2019. "Performance of Complex Visual Tasks Using Simulated Prosthetic Vision via Augmented-Reality Glasses." *Journal of Vision* 19 (13): 22.

Humayun, Mark S., Jessy D. Dorn, Lyndon da Cruz, Gislin Dagnelie, José-Alain Sahel, Paulo E. Stanga, Artur V. Cideciyan, et al. 2012. "Interim Results from the International Trial of Second Sight's Visual Prosthesis." *Ophthalmology* 119 (4): 779–88.

Jepson, Lauren H., Paweł Hottowy, Keith Mathieson, Deborah E. Gunning, Władysław Dąbrowski, Alan M. Litke, and E. J. Chichilnisky. 2014. "Spatially Patterned Electrical Stimulation to Enhance Resolution of Retinal Prostheses." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 34 (14): 4871–81.

Johnson, L. A., J. D. Wander, D. Sarma, D. K. Su, E. E. Fetz, and J. G. Ojemann. 2013. "Direct Electrical Stimulation of the Somatosensory Cortex in Humans Using Electrocorticography Electrodes: A Qualitative and Quantitative Report." *Journal of Neural Engineering* 10 (3): 036021.

Kim, Young Joon, Nora Brackbill, Eleanor Batty, Jinhyung Lee, Catalin Mitelut, William Tong, E. J. Chichilnisky, and Liam Paninski. 2021. "Nonlinear Decoding of Natural Images From Large-Scale Primate Retinal Ganglion Recordings." *Neural Computation* 33 (7): 1719–50.

Kling, A., A. R. Gogliettino, N. P. Shah, E. G. Wu, and N. Brackbill. 2020. "Functional Organization of Midget and Parasol Ganglion Cells in the Human Retina." *bioRxiv*. https://www.biorxiv.org/content/10.1101/2020.08.07.240762v1.abstract.

Lieby, Paulette, Nick Barnes, Chris McCarthy, Nianjun Liu, Hugh Dennett, Janine G. Walker, Viorica Botea, and Adele F. Scott. 2011. "Substituting Depth for Intensity and Real-Time Phosphene Rendering: Visual Navigation under Low Vision Conditions." *Conference Proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference* 2011: 8017–20.

Li, Peter H., Jeffrey L. Gauthier, Max Schiff, Alexander Sher, Daniel Ahn, Greg D. Field, Martin Greschner, Edward M. Callaway, Alan M. Litke, and E. J. Chichilnisky. 2015. "Anatomical Identification of Extracellularly Recorded Cells in Large-Scale Multielectrode Recordings." *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience* 35 (11): 4663–75.

Loudin, J. D., D. M. Simanovskii, K. Vijayraghavan, C. K. Sramek, A. F. Butterwick, P. Huie, G. Y. McLean, and D. V. Palanker. 2007. "Optoelectronic Retinal Prosthesis: System Design and Performance." *Journal of Neural Engineering* 4 (1): S72–84.

Madugula, Sasi, Alex R. Gogliettino, Moosa Zaidi, Gorish Aggarwal, Alexandra Kling, Nishal P. Shah, Ramandeep Vilkhu, et al. 2020. "Focal Electrical Stimulation of Human Retinal Ganglion Cells." https://doi.org/10.1101/2020.08.23.263608.

McCarthy, Chris, Nick Barnes, and Paulette Lieby. 2011. "Ground Surface Segmentation for Navigation with a Low Resolution Visual Prosthesis." *Conference Proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference* 2011: 4457–60.

Mena, Gonzalo E., Lauren E. Grosberg, Sasidhar Madugula, Paweł Hottowy, Alan Litke, John Cunningham, E. J. Chichilnisky, and Liam Paninski. 2017. “Electrical Stimulus Artifact Cancellation and Neural Spike Detection on Large Multi-Electrode Arrays.” *PLoS Computational Biology* 13 (11): e1005842.

Merabet, Lotfi B., and Alvaro Pascual-Leone. 2010. “Neural Reorganization Following Sensory Loss: The Opportunity of Change.” *Nature Reviews. Neuroscience* 11 (1): 44–52.

Palanker, D., Y. Le Mer, Saddek Mohand-Said, M. M. K. Muqit, and J. A. Sahel. 2020. “Photovoltaic Restoration of Central Vision in Atrophic Age-Related Macular Degeneration.” *Ophthalmology*. https://doi.org/10.1016/j.ophtha.2020.02.024.

Parthasarathy, N., E. Batty, W. Falcon, T. Rutten, M. Rajpal, E. J. Chichilnisky, and L. Paninski. 2017. “Neural Networks for Efficient Bayesian Decoding of Natural Images from Retinal Neurons.” *Advances in Neural Information Processing Systems* 30.

Relic, Lucas, Bowen Zhang, Yi-Lin Tuan, and Michael Beyeler. 2022. “Deep Learning-Based Perceptual Stimulus Encoder for Bionic Vision.” *arXiv [cs.CV]*. arXiv. http://arxiv.org/abs/2203.05604.

Richard, Emile, Georges A. Goetz, and E. J. Chichilnisky. 2015. “Recognizing Retinal Ganglion Cells in the Dark.” In *Advances in Neural Information Processing Systems 28*, 2476–84. Curran Associates, Inc.

Rodieck, Robert W. 1998. *The First Steps in Seeing*. Sinauer Associates Incorporated.

Rouger, J., S. Lagleyre, B. Fraysse, S. Deneve, O. Deguine, and P. Barone. 2007. “Evidence That Cochlear-Implanted Deaf Patients Are Better Multisensory Integrators.” *Proceedings of the National Academy of Sciences of the United States of America* 104 (17): 7295–7300.

Ruyter van Steveninck, Jaap de, Umut Güçlü, Richard van Wezel, and Marcel van Gerven. 2022. “End-to-End Optimization of Prosthetic Vision.” *Journal of Vision* 22 (2): 20.

Sekirnjak, Chris, Lauren H. Jepson, Pawel Hottowy, Alexander Sher, Wladyslaw Dabrowski, A. M. Litke, and E. J. Chichilnisky. 2011. “Changes in Physiological Properties of Rat Ganglion Cells during Retinal Degeneration.” *Journal of Neurophysiology* 105 (5): 2560–71.

Shah, Nishal P., and E. J. Chichilnisky. 2020. “Computational Challenges and Opportunities for a Bi-Directional Artificial Retina.” *Journal of Neural Engineering* 17 (5): 055002.

Soto, Florentina, Jen-Chun Hsiang, Rithwick Rajagopal, Kisha Piggott, George J. Harocopos, Steven M. Couch, Philip Custer, Josh L. Morgan, and Daniel Kerschensteiner. 2020. “Efficient Coding by Midget and Parasol Ganglion Cells in the Human Retina.” *Neuron* 107 (4): 656–66.e5.

Stingl, Katarina, Karl Ulrich Bartz-Schmidt, Dorothea Besch, Angelika Braun, Anna Bruckmann, Florian Gekeler, Udo Greppmaier, et al. 2013. “Artificial Vision with Wirelessly Powered Subretinal Electronic Implant Alpha-IMS.” *Proceedings. Biological Sciences / The Royal Society* 280 (1757): 20130077.

Tandon, Pulkit, Nandita Bhaskhar, Nishal Shah, Sasi Madugula, Lauren Grosberg, Victoria H. Fan, Pawel Hottowy, et al. 2021. “Automatic Identification of Axon Bundle Activation for Epiretinal Prosthesis.” *IEEE Transactions on Neural Systems and Rehabilitation Engineering: A Publication of the IEEE Engineering in Medicine and Biology Society* 29 (December): 2496–2502.

Trenholm, Stuart, and Gautam B. Awatramani. 2015. “Origins of Spontaneous Activity in the Degenerating Retina.” *Frontiers in Cellular Neuroscience* 9 (July): 277.

Vergnieux, Victor, Marc J-M Macé, and Christophe Jouffrais. 2017. “Simplification of Visual Rendering in Simulated Prosthetic Vision Facilitates Navigation.” *Artificial Organs* 41 (9): 852–61.

Vilkhu, Ramandeep S., Sasidhar S. Madugula, Lauren E. Grosberg, Alex R. Gogliettino, Pawel

Hottowy, Wladyslaw Dabrowski, Alexander Sher, Alan M. Litke, Subhasish Mitra, and E. J. Chichilnisky. 2021. “Spatially Patterned Bi-Electrode Epiretinal Stimulation for Axon Avoidance at Cellular Resolution.” *Journal of Neural Engineering* 18 (6). https://doi.org/10.1088/1741-2552/ac3450.

Wandell, Brian A. 1995. *Foundations of Vision*. Sinauer Associates, Incorporated.

Wang, Zhou, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P. Simoncelli. 2004. “Image Quality Assessment: From Error Visibility to Structural Similarity.” *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society* 13 (4): 600–612.

Warland, David K., Pamela Reinagel, and Markus Meister. 1997. “Decoding Visual Information From a Population of Retinal Ganglion Cells.” *Journal of Neurophysiology*. https://doi.org/10.1152/jn.1997.78.5.2336.

Wu, Eric Gene, Alexander Sher, Alan M. Litke, Eero P. Simoncelli, and E. J. Chichilnisky. 2022. “Maximum a Posteriori Natural Scene Reconstruction from Retinal Ganglion Cells with Deep Denoiser Priors.” *bioRxiv*. https://doi.org/10.1101/2022.05.19.492737.

Yarbus, Alfred L. 1967. “Eye Movements and Vision.” https://doi.org/10.1007/978-1-4899-5379-7.

Zaidi, Moosa, Gorish Aggarwal, Nishal P. Shah, Orren Karniol-Tambour, Georges Goetz, Sasi Madugula, Alex R. Gogliettino, et al. 2022. “Inferring Light Responses of Primate Retinal Ganglion Cells Using Intrinsic Electrical Signatures.” *bioRxiv*. https://doi.org/10.1101/2022.05.29.493858.